ARMY RESEARCH LABORATORY

Weather Research and Forecasting (WRF) Results over New Mexico

by Barbara Sauter

ARL-TR-3715 January 2006

NOTICES

Disclaimers

Army Research Laboratory

White Sands Missile Range, NM 88002-5501

ARL-TR-3715 **January 2006**

Weather Research and Forecasting (WRF) Results over New Mexico

Barbara Sauter
Computational and Information Sciences Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	January 2006
2. REPORT TYPE	Final
3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE	Weather Research and Forecasting (WRF) Results over New Mexico
5a. CONTRACT NUMBER	
5b. GRANT NUMBER	
5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)	Barbara Sauter
5d. PROJECT NUMBER	
5e. TASK NUMBER	
5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	U.S. Army Research Laboratory Computational and Information Sciences Directorate Battlefield Environment Division (ATTN: AMSRD-ARL-CI-EM) White Sands Missile Range, NM 88002-5501
8. PERFORMING ORGANIZATION REPORT NUMBER	ARL-TR-3715
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	U.S. Army Research Laboratory 2800 Powder Mill Road Adelphi, MD 20783-1145
10. SPONSOR/MONITOR'S ACRONYM(S)	ARL
11. SPONSOR/MONITOR'S REPORT NUMBER(S)	ARL-TR-3715
12. DISTRIBUTION/AVAILABILITY STATEMENT	Approved for public release; distribution is unlimited.
13. SUPPLEMENTARY NOTES	

14. ABSTRACT

Researchers from the National Center for Atmospheric Research (NCAR) and other organizations are developing the Weather Research and Forecasting (WRF) model with the goal of meeting the needs of both the meteorological research and operational weather forecast communities. This report documents the results found running the WRF version 2.0.3.1 in a nested mode centered over southern New Mexico. The WRF model, initialized with North American Mesoscale (NAM) model output, comprised three domains based on grid spacing intervals of 18 km, 6 km, and 2 km. Model output is compared to surface observations and upper-air soundings for all three domains.

15. SUBJECT TERMS

Weather Research and Forecasting; WRF; Forecast Errors

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	SAR	46	Barbara Sauter

19b. TELEPHONE NUMBER (*Include area code*)
(505) 678-2840

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

Preface

Model evaluations are highly dependent on the availability and quality of data used for comparisons with model output. This study concentrates primarily on the Weather Research and Forecasting (WRF) model basic weather parameters over White Sands Missile Range, NM, where surface observations based on Army Surface Atmospheric Measuring System (SAMS) data are considered sufficiently reliable and accurate. In addition, surface observations available through the National Oceanographic and Atmospheric Administration Forecast Systems Laboratory's Meteorological Assimilation Data Ingest System (MADIS) provide supplemental comparisons for the two larger domains. Since MADIS data are not centrally managed, quality control of the data is not certain. In order to ensure proper comparisons, results are also provided for the WRF output based on grid spacings of 18 km, 6 km, and 2 km, using only the data from the inner-most domain, so each forecast is evaluated against the exact same observations for those cropped cases. Vertical output is analyzed only at locations with National Weather Service upper-air soundings. The 18-km outer domain covers 3 upper-air locations, while the 2 subnest domains only include 1 upper-air location.

The author acknowledges the valuable advice and assistance received from Dr. Teizi Henmi in setting up the WRF files and from Robert Flanigan in data retrieval and handling for the surface observations.

Summary

The Weather Research and Forecasting (WRF) model was run in three nests centered over southern New Mexico. The grid spacing for the three nests was 18 km, 6 km, and 2 km. The dates of the forecasts ranged between March 28 and May 15, 2005, encompassing 45 days with the model initialized at 0000 Zulu (00Z) time and 34 days with the model initialized at 1200 Zulu (12Z). Output of basic surface weather parameters was generated hourly out to a 12-h forecast, along with a 0-h and 12-h upper-air forecast for each initialization time. This report documents the results seen comparing the WRF model output of temperature, dew-point temperature, wind speed, and wind direction to observed data. The surface observation data were based on values from the National Oceanographic and Atmospheric Administration Forecast Systems Laboratory's Meteorological Assimilation Data Ingest System (MADIS) stations and the Surface Atmospheric Measuring System (SAMS) network at the U.S. Army's White Sands Missile Range, NM. Upper-air observations were used from the National Weather Service sounding stations at Tucson, AZ; Albuquerque, NM; and El Paso, TX.

Traditional summary error statistics, including mean and absolute errors, are presented for all the stations in the three nested domains. In addition, the 18-km and 6-km model outputs were cropped to the 2-km model output domain to provide comparisons of the various model grid spacing results based on the same stations. Some upper-air error statistics are provided for the single height of 50 kPa, along with subjective discussions on the quality of the entire upper-air profiles.

INTENTIONALLY LEFT BLANK.

1. Introduction

Over the past several decades, global weather forecast models have been the basis for improving weather forecasts. Mesoscale models can provide further details for regional areas of interest. Civilian and military communities are expected to adopt the Weather Research and Forecasting (WRF) mesoscale model as a primary tool for both research and operational weather forecasting as the WRF continues to be developed and improved by the National Center for Atmospheric Research (NCAR) and other organizations (*1*). A nested version of the WRF model became available in late 2004. This study documents the evaluation of three nested WRF domains, comparing model output to surface and upper-air observations of the basic weather parameters of temperature, dew-point temperature, and wind.

The Army is particularly interested in model accuracy over areas of complex terrain. The location chosen for this study contains varying terrain, with the primary analyses focusing on an area in southern New Mexico.

2. Methodology

2.1 Forecast Model Runs

The WRF model runs were initialized with 40-km North American Mesoscale (NAM) model output data (*2, 3*). The three nested WRF domains are described in table 1.

Table 1. WRF domain information.

Domain	Center	Number of Points		Grid Spacing[a] (km)	Vertical Levels
		X×Y	Total		
d01	33.5° N. 107.5° W.	50×40	2,000	18	31
d02	32.6° N. 106.8° W.	67×61	4,087	6	31
d03	32.8° N. 107.1° W.	103×118	12,154	2	31

[a] Distance between points.

The outer domain (d01) encompasses most of New Mexico and eastern Arizona, as well as extending into a small portion of northern Mexico and southwest Texas (figure 1). This domain includes numerous peaks and valleys, with terrain elevations ranging from less than 500 m to more than 3,000 m.

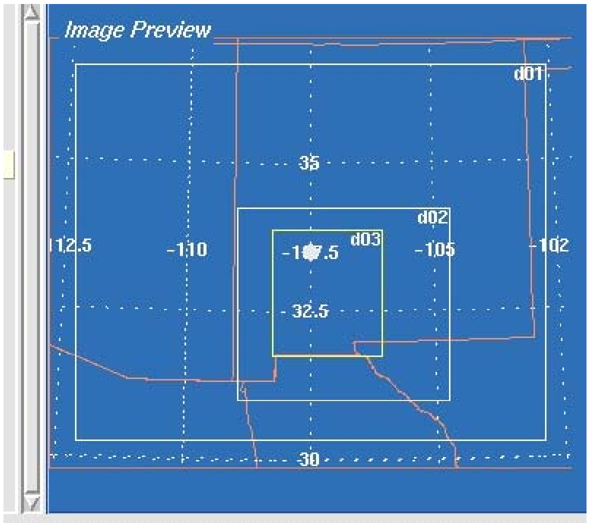


Figure 1. WRF nested domains, labeled d01, d02, and d03 in the upper right corner of each domain box.

Model options include the following choices:

- microphysics option 2 Lin et al. scheme
- long-wave radiation option 1 Rapid Radiation Transfer Model (RRTM) scheme
- short-wave radiation option 1 Dudhia scheme
- surface-layer option 1 Monin-Obukhov scheme
- land-surface option 2 Noah land-surface model
- boundary layer option 1 Yonsei University (YSU) scheme
- cumulus option 1 Kain-Fritsch (new Eta) scheme

The model generated hourly output up to 12 h. Springtime runs were initialized at 0000 Zulu (00Z) time for 45 days between March 28 and May 15, 2005. In addition, runs were initialized at 1200 Zulu (12Z) for each of the 34 days from April 12 through May 15, 2005. The 00Z model runs produced forecasts for 5 p.m. through 5 a.m. Mountain Standard Time (MST), while the 12Z runs produced forecasts for 5 a.m. through 5 p.m. MST.

2.2 Surface Validation Stations

Surface observations used for comparison with model forecasts were obtained from two sources: the National Oceanographic and Atmospheric Administration Forecast Systems Laboratory's Meteorological Assimilation Data Ingest System (MADIS) and the Surface Atmospheric Measuring System (SAMS) network at the U.S. Army's White Sands Missile Range, NM (*4*). The MADIS observations are provided from a number of different sources and may not strictly satisfy prevalent standards for sensor calibration, height, or observation times. The SAMS data are considered to be more reliable relative to those considerations. The same stations were not always available for each hour of a model's run, and many changes in station availability occurred between the different days used in this study. Approximately 70 MADIS and 20 SAMS surface stations reported observations multiple times throughout the 45-day study period. Figure 2 shows the range of elevation heights for the surface stations.

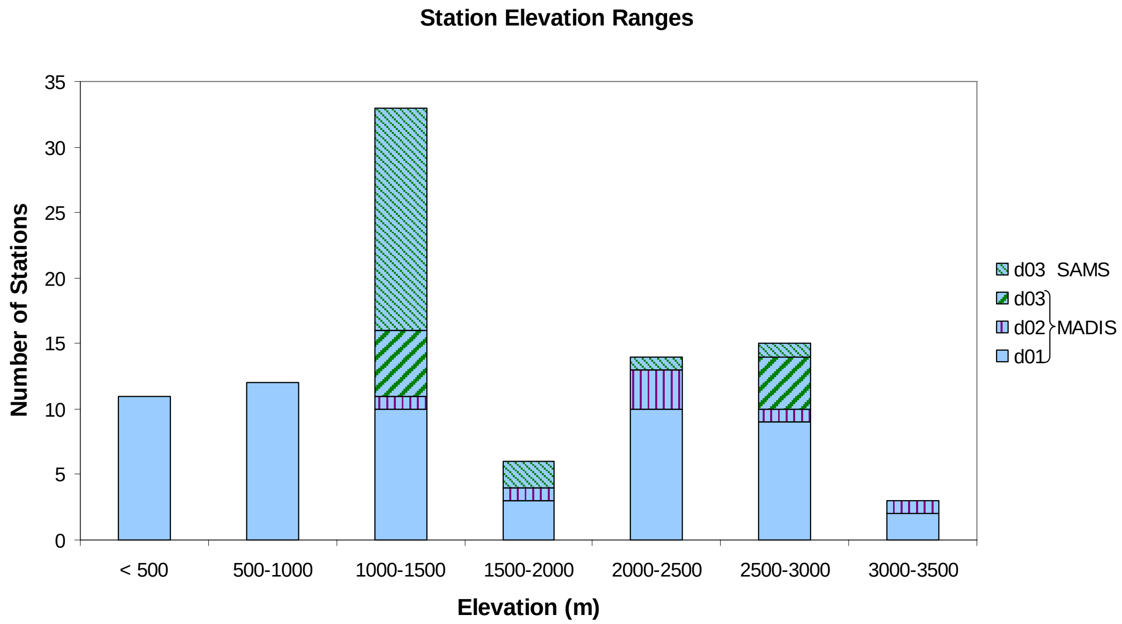


Figure 2. The number of surface stations at various altitude ranges.

NOTE: Domain 1 includes all stations listed as d01, d02, and d03. Domain 2 includes all stations listed as d02 and d03.

For an individual forecast time, typically about 50 surface stations were available from MADIS within domain 1, although a third of those stations might only report temperature with no dew-point temperature or wind information. Out of those approximately 50 stations, about 10 were also in the area covered by domain 2, and 3 were in the innermost domain 3. Approximately 18 SAMS observations were available, located on the White Sands Missile Range, NM, over the eastern half of domain 3 (figure 3).

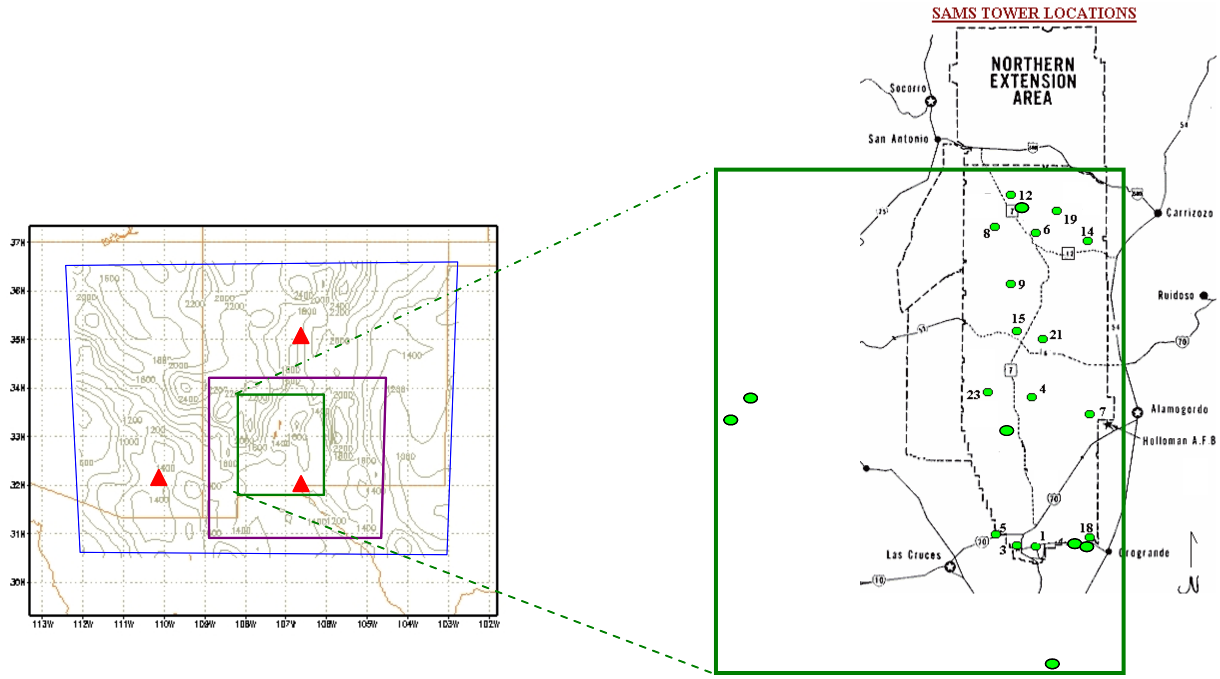


Figure 3. The upper-air sounding locations for the nested domains (highlighted with red triangles) and the surface observation locations for domain 3 only (indicated by the green ovals).

2.3 Upper-Air Validation Stations

Three National Weather Service upper-air sounding observation sites are located within the boundaries of domain 1 (figure 3). Information on these stations is provided in table 2.

Table 2. Upper-air sounding stations.

Station Name	Identifier	Location	Elevation (m)
El Paso, TX	KELP	31.8º N. 106.4º W.	1,194
Tucson, AZ	KTUS	32.1º N. 110.9º W.	779
Albuquerque, NM	KABQ	35.1º N. 106.6º W.	1,620

The El Paso, TX, upper-air station data may be used for all 3 domain model output validations based on grid-point spacings of 18 km, 6 km, and 2 km. The other 2 sounding sites are only within the largest domain 1, based on an 18-km grid-point spacing.

3. Results

The primary purpose of this study was to document the general performance of the WRF model over one area and one timeframe. Since the forecasts from the 3 nested domains were all available, many of the results shown compare the errors between the 18-km, 6-km, and 2-km grid-space nests. The problem of seeing higher standard error measures associated with finer resolution grid spacing is well known (*5*). This problem is not expected to be as great for the basic parameters of temperature, dew-point temperature, wind speed, and wind direction presented here as it frequently is for small-scale phenomena, such as isolated rainfall.

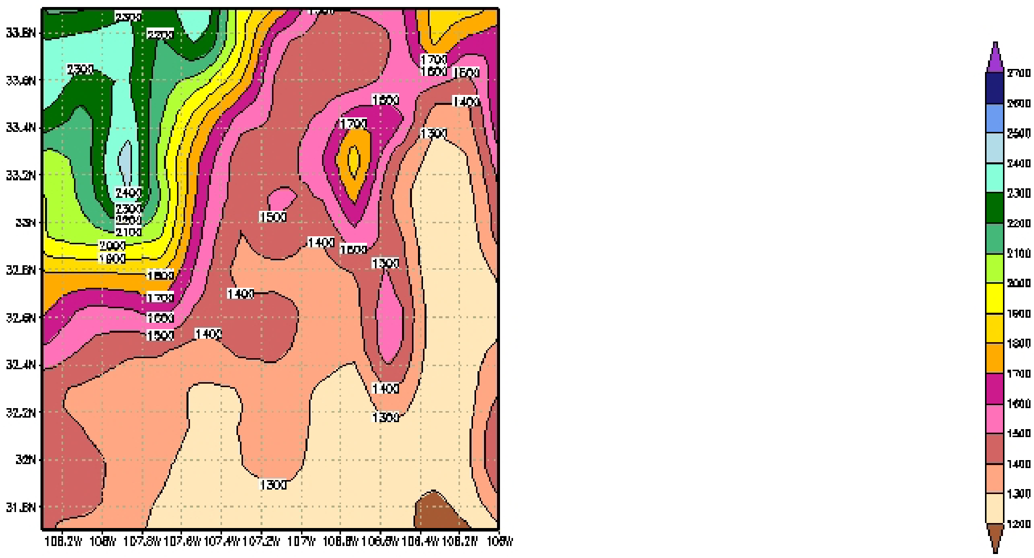
3.1 Surface Forecast Evaluations

Mean and absolute error statistics were calculated based on all the surface observations available within each domain. In addition, to ensure comparisons between the various grid-spacing forecasts were based on the exact same observation quality and forecast difficulty, error statistics were calculated for the 18-km and 6-km model runs with the output cropped to the area within the 2-km model run. Information relating to these subsets of error statistics is shown in table 3.

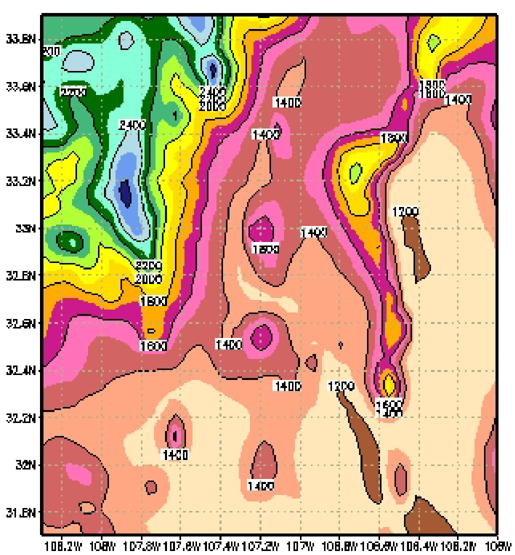
Table 3. Identification of subsets of model output error statistics.

Identifier	Model Grid Spacing (km)	Area Coverage (km)
d01	18	domain 1 882×702
d02	6	domain 2 396×360
d03	2	domain 3 204×234
d01c3	18	domain 3 204×234
d02c3	6	domain 3 204×234

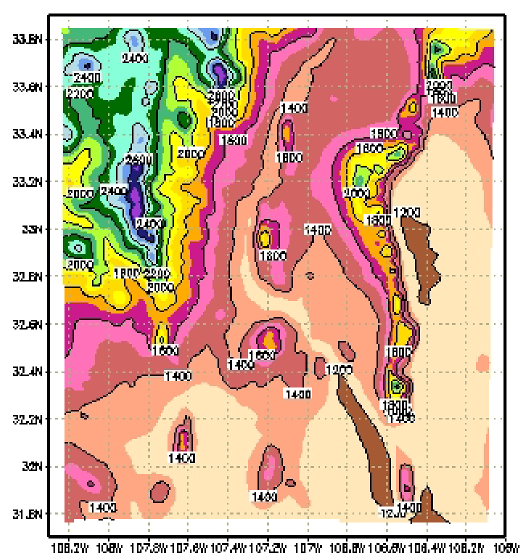
A primary advantage of the finer-scale 2-km WRF model run is the greater detail known about the terrain elevation at specific sites. Figure 4 illustrates this detail, highlighting the smooth terrain associated with the 18-km model output cropped to the domain 3 area (figure 4a), becoming more detailed and picking up additional variation and higher peaks as the model grid spacing increases to 6 km (figure 4b) and to 2 km (figure 4c).



(a)



(b)



(c)

Figure 4. Surface elevation heights associated with WRF output for (a) domain 1 (18-km grid spacing), (b) domain 2 (6-km grid spacing), and (c) domain 3 (2-km grid spacing).

NOTE: All domains are cropped to the area covered by domain 3.

3.1.1 Temperature Errors

The box-whisker plot shown in figure 5 shows the wide range of temperature values experienced throughout the various days and locations in this study for the largest domain, with the forecast data based on the 18-km grid-space forecasts (d01). For the specified time, one-quarter of the values fell within each of the temperature ranges indicated by 1) the bottom whisker (or cross) and the bottom of the box, 2) the lower box, 3) the upper box, and 4) the top of the upper box to the highest whisker (or cross). Values that are outside the lower or upper limits of the boxes by more than 1.5 times the vertical extent of the boxes are indicated with a cross.

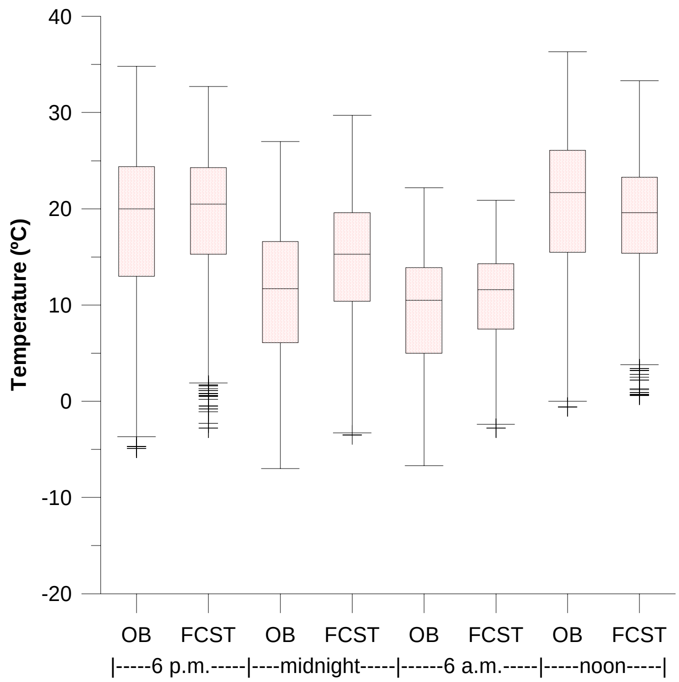


Figure 5. Quartiles of temperature values observed and forecast within domain 1 for all the stations and dates.

NOTE: OB = observations and FCST = forecast.

The 6 p.m. and midnight MST data incorporate approximately 2,500 data points based on the 45 days used for the 00Z model runs, while the 6 a.m. and noon MST data include approximately 2,000 data points based on the 34 days used for the 12Z model runs. The 6 p.m. and 6 a.m. MST results are both from 1-h forecasts and the midnight and noon MST results are both based on 7-h forecasts. These times all show a predominant warm temperature bias, except for the noon MST forecasts. The spread of forecast values is somewhat less than the spread of the observations, but the range of values depicted is quite large for both the forecast and observed data.

The average mean errors for all the temperature forecasts are shown separately for the 00Z and 12Z model runs in figure 6. The average absolute errors are shown in figure 7. The evaluation times for the 00Z model runs include each hour from 0100 Zulu (01Z) through 12Z (6 p.m. through 5 a.m. MST), while the 12Z model runs include each hour from 1300 Zulu (13Z) through 00Z (6 a.m. through 5 p.m. MST). The most pronounced difference is between the large 3–4 ºC warm bias in the 00Z forecasts and the smaller approximately 1 ºC cool bias in the 12Z forecasts. The 00Z bias is most pronounced in the 18-km results over the innermost domain (d01c3), which is associated with the highest absolute error of 4.4 ºC. The 00Z WRF errors generated by the 6-km grid spacing nest cropped to the inner domain (d02c3) show a third of a degree improvement over the 18-km model run, and the 2-km innermost nest shaves off a half of a degree error from the 18-km model run. The improvements are more significant in the 12Z WRF runs, with over a half degree improvement from the 18-km results to the 6-km results and almost a full degree improvement from the 18-km results to the 2-km results, out of the smaller 3.2 ºC average 18-km absolute error.

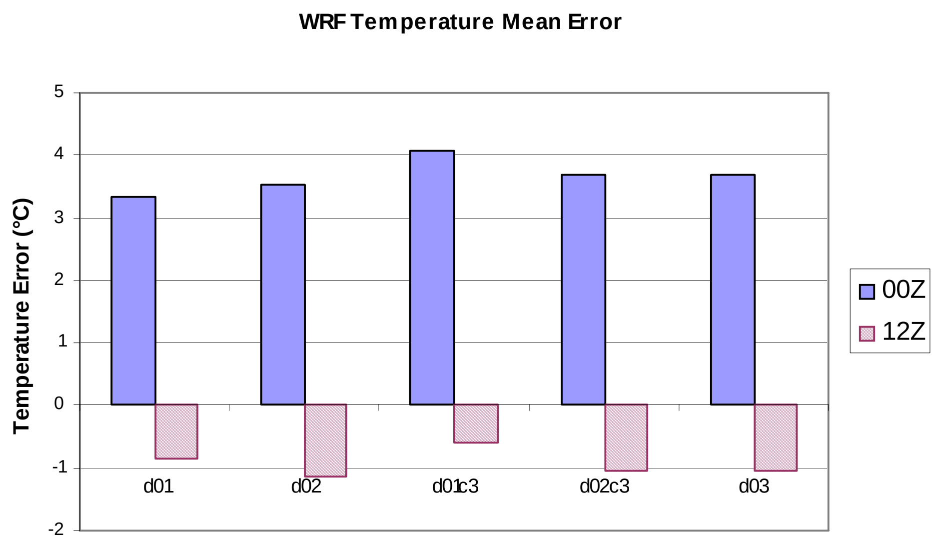


Figure 6. Average mean temperature errors for all stations and times for the 00Z and 12Z WRF model runs.

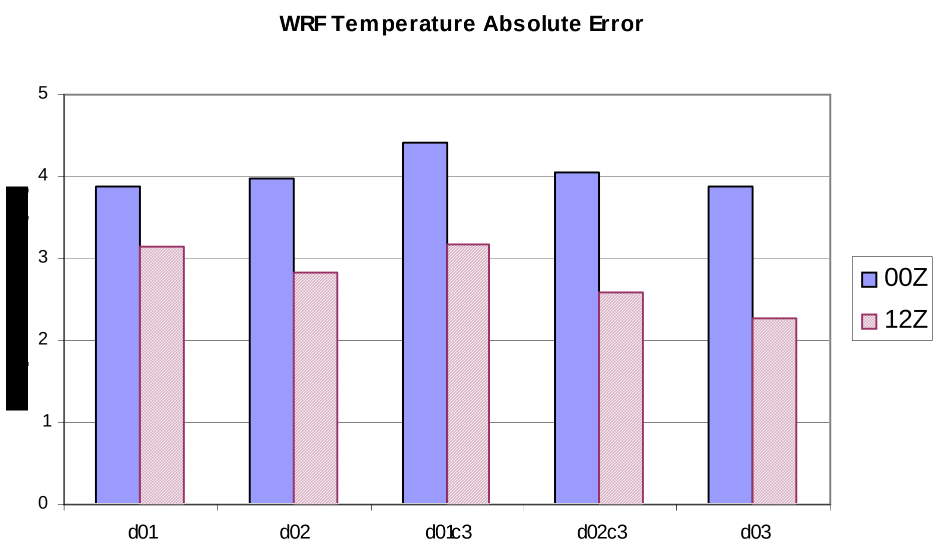


Figure 7. Average absolute temperature errors for all stations and times for the 00Z and 12Z WRF model runs.

Although the inner 2-km grid produced smaller temperature errors than generated by the outer nests with larger grid spacings, the temperature errors are still larger than desired, particularly for the model runs initialized at 00Z. A previous study suggested a possible correlation between forecast errors and station elevation (*6*). Error statistics were not calculated for individual stations for this study, but it is interesting to note the results of a temperature forecast for one specific time considering the elevations of the stations within the innermost domain. Figure 8 shows these stations on the horizontal axis arranged from the lowest to highest elevation, as indicated by the vertical axis on the right.

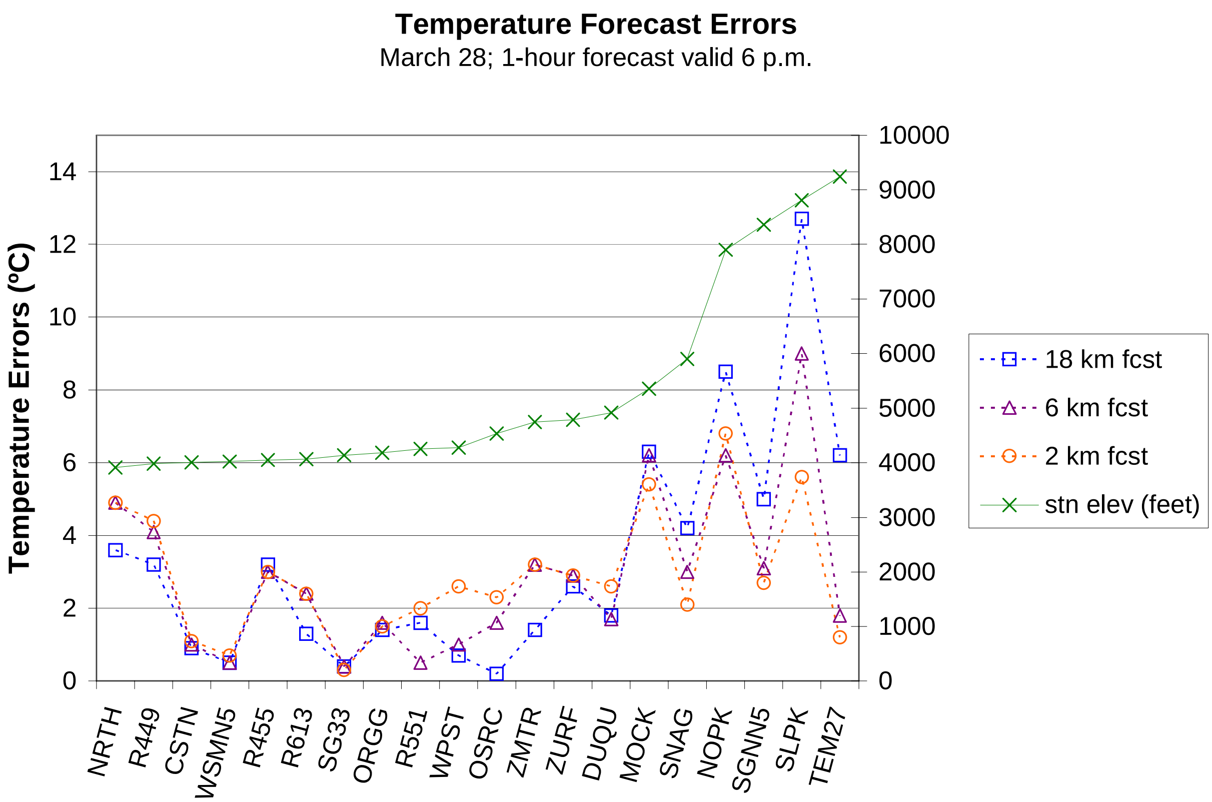


Figure 8. Temperature forecast errors for a single date and time for stations in the innermost domain, with the station elevation indicated by the right vertical axis.

The majority of SAMS sites are at an elevation of approximately 4,000 ft. These sites show absolute temperature errors between 0 and 5 ºC on March 28, 2005 at 6 p.m. MST. The 2-km forecasts are generally the same or even worse than the 6-km and 18-km forecasts for these stations. On the other hand, several of the stations with elevations above 5,000 ft experienced significantly higher forecast errors for the 18-km nest, with better results for the 6-km and 2-km nests. The stations labeled SGNN5 and TEM27 are MADIS stations at the far western edge of domain 3 (figure 3), where the surrounding terrain is also relatively high. In contrast, the SAMS stations labeled NOPK and SLPK are situated on more isolated peaks extending out of the desert floor, and those two stations show the highest 18-km forecast error. The two stations with the highest elevation both show great improvements in the 2-km forecast as compared to the 18-km forecast.

3.1.2 Dew-point Temperature Errors

The dew-point temperatures observed throughout this study also cover a substantial range (see figure 9).

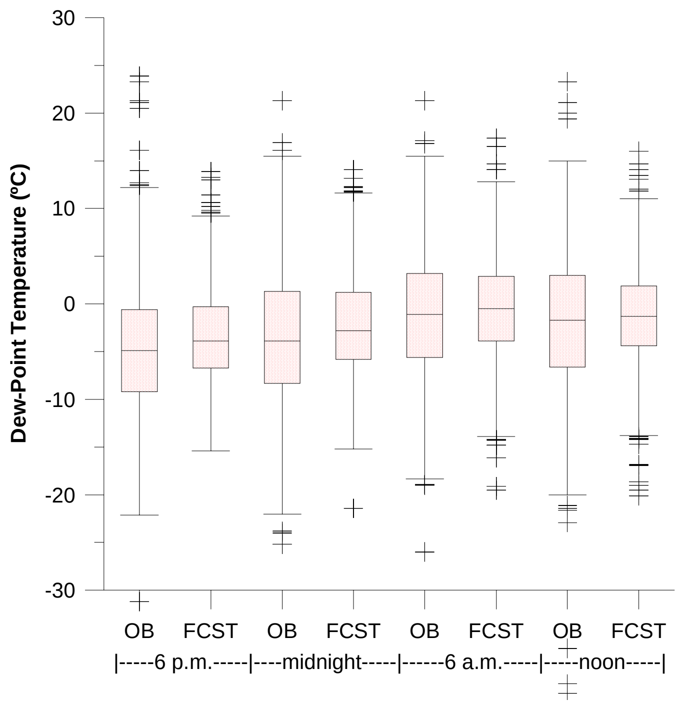


Figure 9. Quartiles of dew-point temperature values observed and forecast within domain 1 for all the stations and dates.

NOTE: OB = observations and FCST = forecast.

There are approximately 500 fewer observations for dew-point temperature than for temperature, with close to 2,000 observations contained in the 6 p.m. and midnight MST data and 1,500 observations in the 6 a.m. and noon MST data. The dew-point temperature spread within the Inner Quartile Range (IQR) (middle 50% within the boxes) is not quite as widely dispersed as the temperature IQR, and the outliers seem to be more anomalous from the prevailing data. The cross marks in figure 9 indicating values below the *x*-axis are likely to be erroneous observations. The forecasts show a warm bias with less dispersion than seen in the observations for each of the four times.

The dew-point temperature forecasts based on the 00Z model runs exhibit a warm bias, but the bias is not as strong as one in the temperature forecasts (see figure 10). Unlike the 12Z temperature results, the 12Z dew-point temperature forecasts also contained a warm bias, but to a lesser extent than the 00Z model output. The bias was least pronounced when considering all the stations within the largest domain 1 based on the 18-km model runs. There are no significant differences in the dew-point temperature forecast errors generated by the 18-km, 6-km, and 2-km nests cropped to the innermost domain 3.

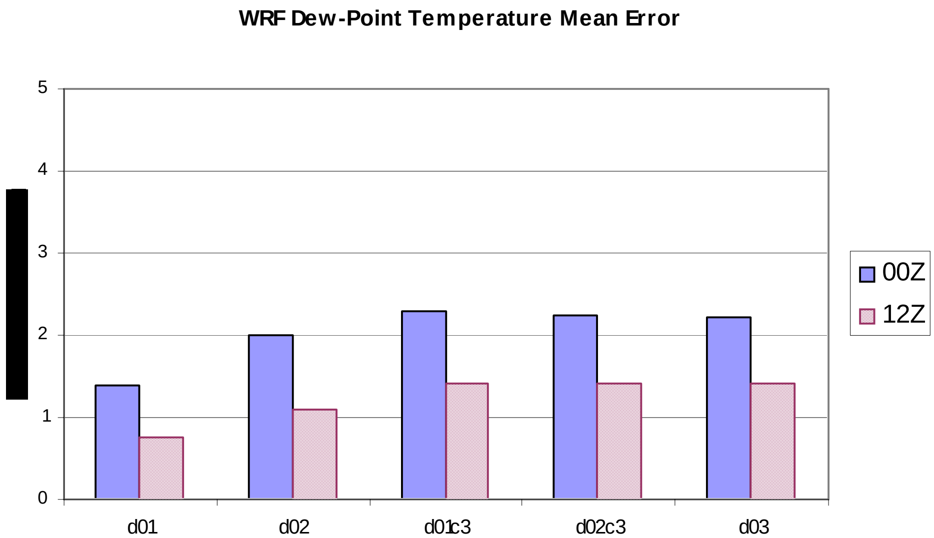


Figure 10. The average mean dew-point temperature errors for all stations and times for the 00Z and 12Z WRF model runs.

Although the 18-km forecast to observation comparisons using all the stations in domain 1 contained the lowest bias, they still had slightly higher average absolute dew-point temperature errors than the other categories (see figure 11). All these errors for the three different nests cropped to domain 3 were close to 3.3 ºC for the 00Z runs and 3.0 ºC for the 12Z runs.

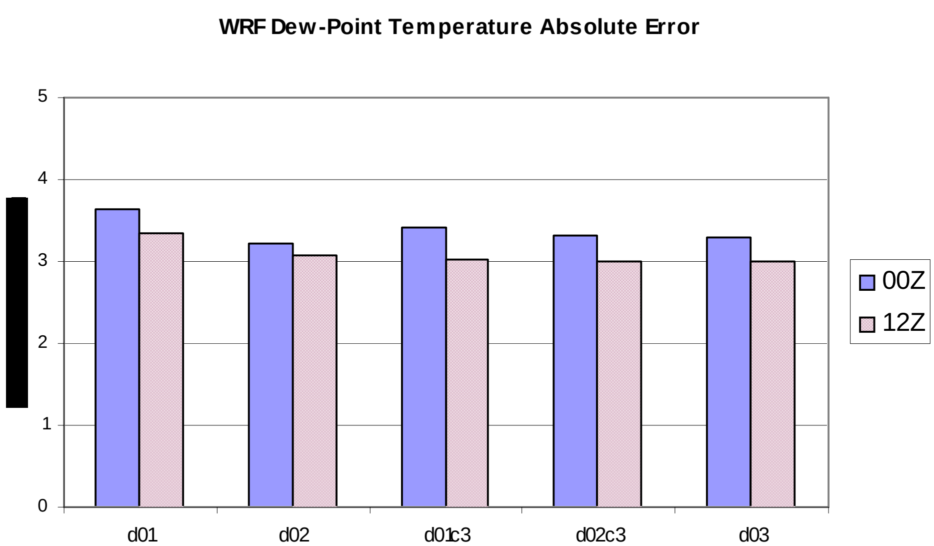


Figure 11. The average absolute dew-point temperature errors for all stations and times for the 00Z and 12Z WRF model runs.

The high-elevation SAMS sites associated with high temperature forecast errors for the 6 p.m. MST forecast on March 28, 2005, did not reflect similar dew-point temperature errors greater than the lower stations (figure 12). However, the higher elevation stations were again most likely to show an improvement in the 2-km forecast when compared to the 18-km forecast. The high-elevation MADIS sites did not report dew-point temperatures for that time.

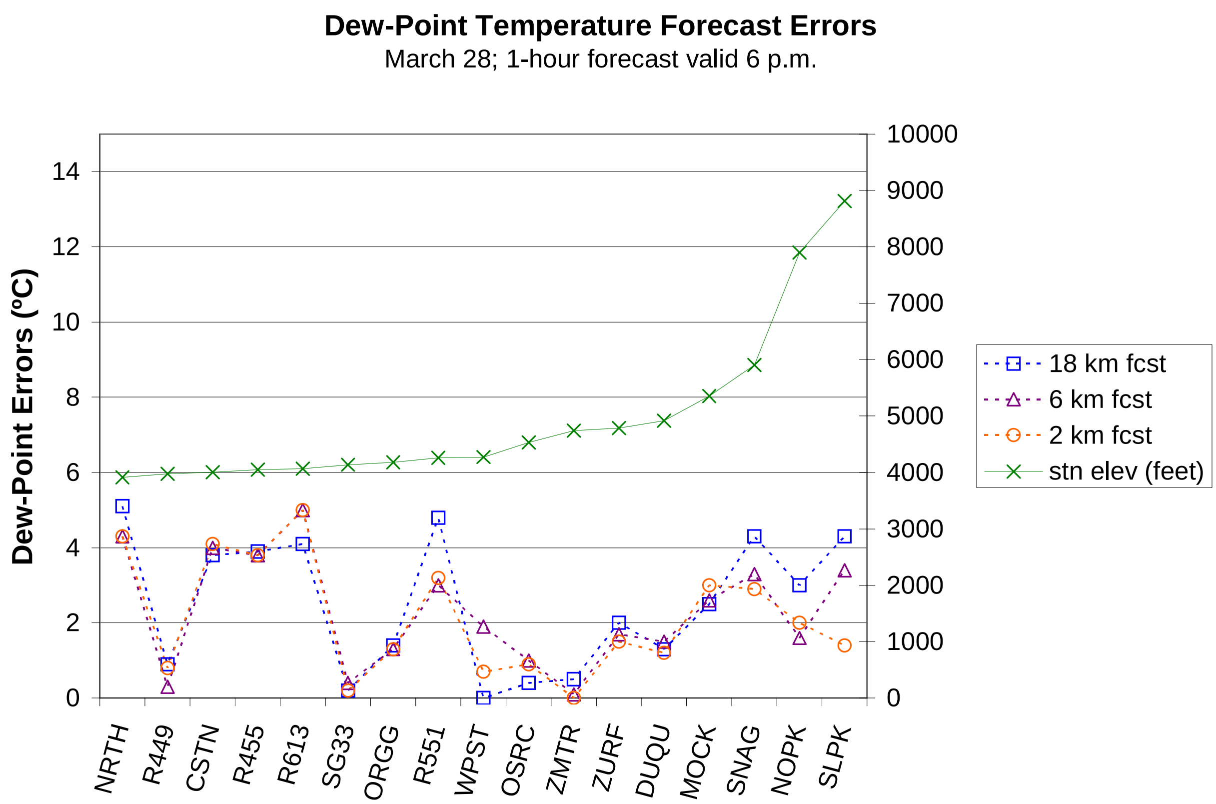


Figure 12. Dew-point temperature forecast errors for a single date and time for stations in the innermost domain, with the station elevation indicated by the right vertical axis.

3.1.3 Wind Speed Errors

Figure 13 depicts the ranges of wind speeds observed throughout this study. Again there are approximately 500 fewer stations reporting wind speed than those reporting dew-point temperature, resulting in close to 1,500 observations for the 00Z model run times (6 p.m. and midnight MST) and 1,000 observations for the 12Z model run times (6 a.m. and noon MST). The IQR shows predominant wind speeds of 3–5 m/s at the 6 p.m. and noon MST observation times, with weaker winds at midnight and 6 a.m. MST, as expected. Although the wind speed forecasts show a bias to be too strong, they do not encompass any of the few extremely high winds observed above 15 m/s.

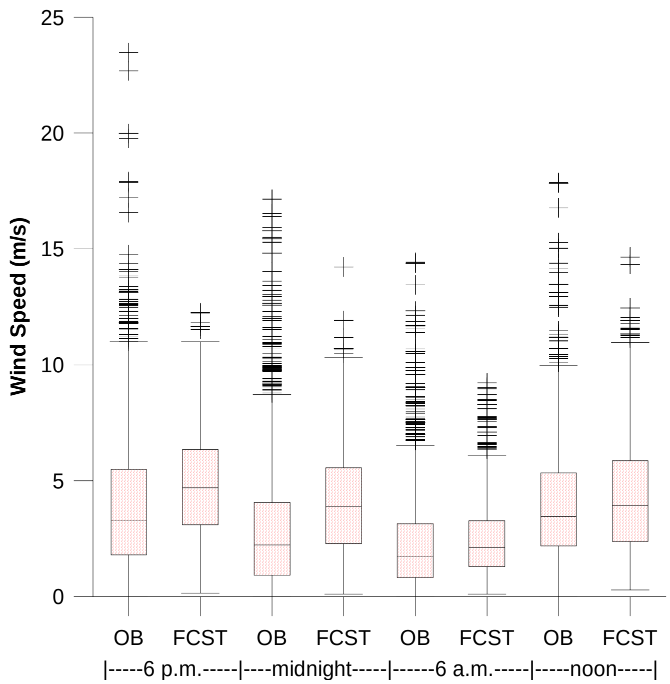


Figure 13. Quartiles of wind speed values observed and forecast within domain 1 for all the stations and dates.

NOTE: OB = observations and FCST = forecast.

The wind speed mean errors reflect a fair amount of variability between the different model domains and run times (figure 14), while the average absolute errors are much more consistent across all the categories (figure 15). The *y*-axis for each plot encompasses 3 m/s and the differences in the bar chart heights are directly comparable.

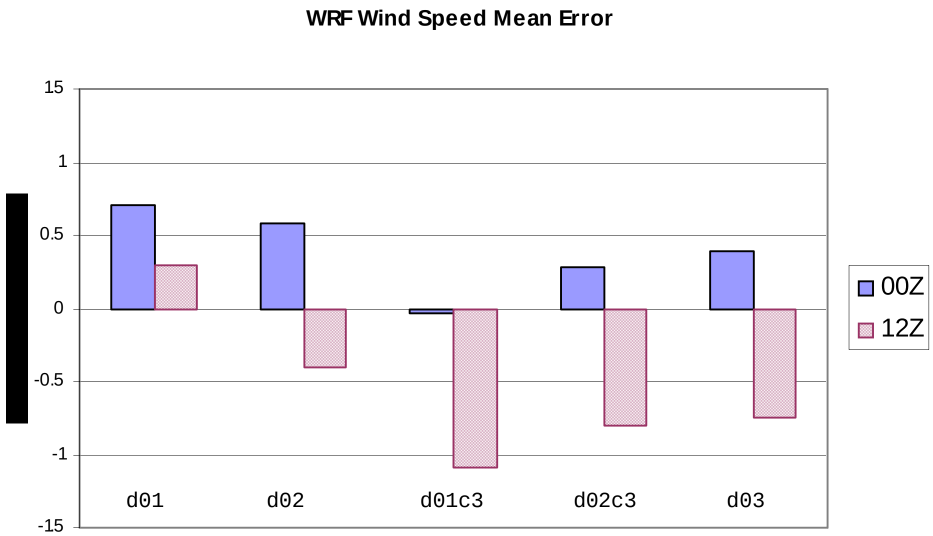


Figure 14. Average mean wind speed errors for all stations and times for 00Z and 12Z WRF model runs.

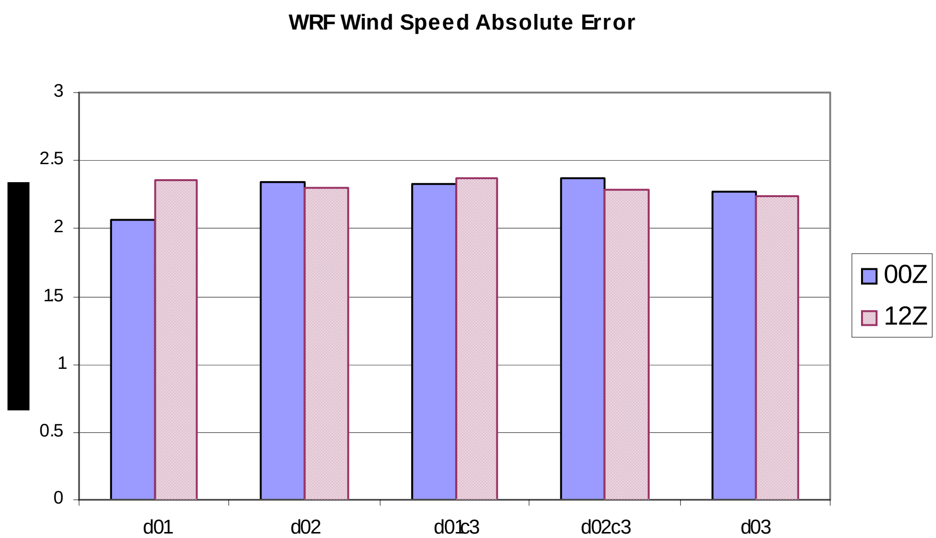


Figure 15. Average absolute wind speed errors for all stations and times for the 00Z and 12Z WRF model runs.

The apparent bias for wind speed forecasts to be too strong based on the 00Z model run may be exaggerated in this study, since the cup anemometers might not start registering wind speeds accurately until a minimum speed has been reached. In fact, 10% of the 6 p.m. and midnight MST observations produced a wind speed of 0 m/s, with the lowest wind speed reported of 0.4 m/s. On the other hand, the forecasts did not include any totally calm wind speeds. The 6 a.m. and noon MST observations included in the 12Z model run comparisons also incorporated observed wind speeds of 0 m/s about 10% of the time, so the model bias shown for domains 2 and 3 with wind speeds forecasted too low may not reflect as large of a bias as actually exists.

The average absolute wind speed error near 2 m/s for the 00Z 18-km model runs for all of domain 1 is slightly lower than the other categories, which are all closer to 2.3 m/s.

3.1.4 Wind Direction Errors

The predominant wind directions experienced during this study are displayed in figure 16 in a series of wind roses for both the observed and the forecasted wind directions. The number of instances of wind blowing from a particular direction is indicated by the length of the wedge extending toward that direction. Since the purpose is to portray the relative percentage of time the wind is from various directions at a specific time, the *x*-axis showing the number of occurrences has been kept consistent within the side-by-side observation and forecast rose plots, but has been allowed to vary across the different times. The number of observations is less than those reported for wind speeds depending on how many calm wind speeds were observed with no corresponding wind direction, leading to 1,400 data points at 6 p.m. MST, 1,200 at midnight MST, 900 at 6 a.m. MST, and 1,000 at noon MST.

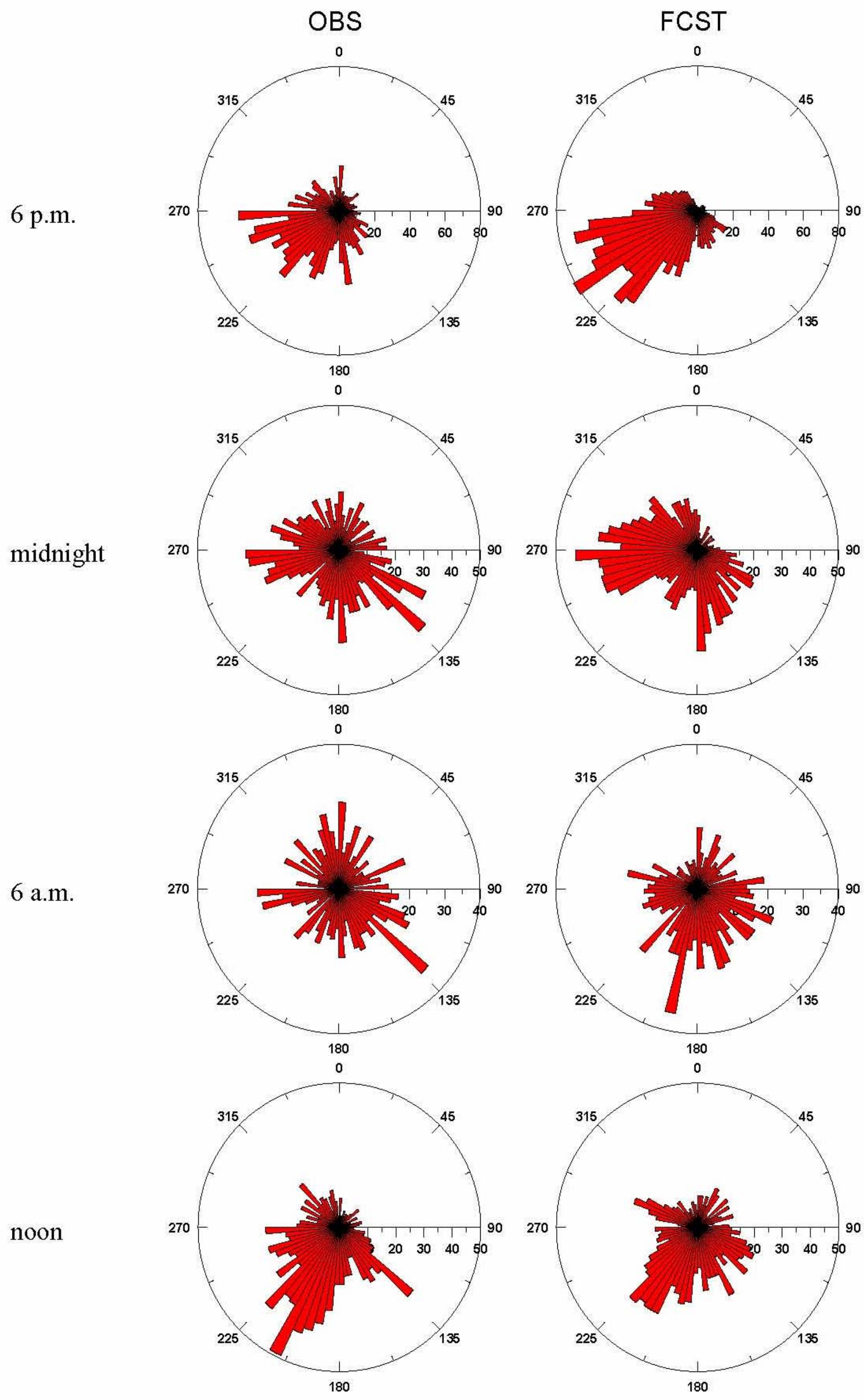


Figure 16. The wind directions experienced during this study for all the stations and dates within domain 1.

NOTE: The length of the red wedges indicates the number of instances the wind was observed or forecast to be blowing from the direction (in degrees) that the wedge extends toward. OBS = observations and FCST = forecast.

The average mean wind direction error has not been calculated, but the wind rose plots in figure 16 don't appear to reflect a clear bias for winds to be forecasted from a more clockwise or counterclockwise direction than the observed values. Even though the observations seem somewhat more likely to be reported from one of the eight main compass points than from a direction not divisible by 45, sufficient observations fall within the other 5º bin wedges that this apparent reporting preference probably does not contribute significantly to wind direction forecast errors. The average absolute wind direction errors are shown in figure 17. Since wind direction observations associated with low wind speeds were not discarded, light and variable winds may contribute significantly to the wind direction errors.

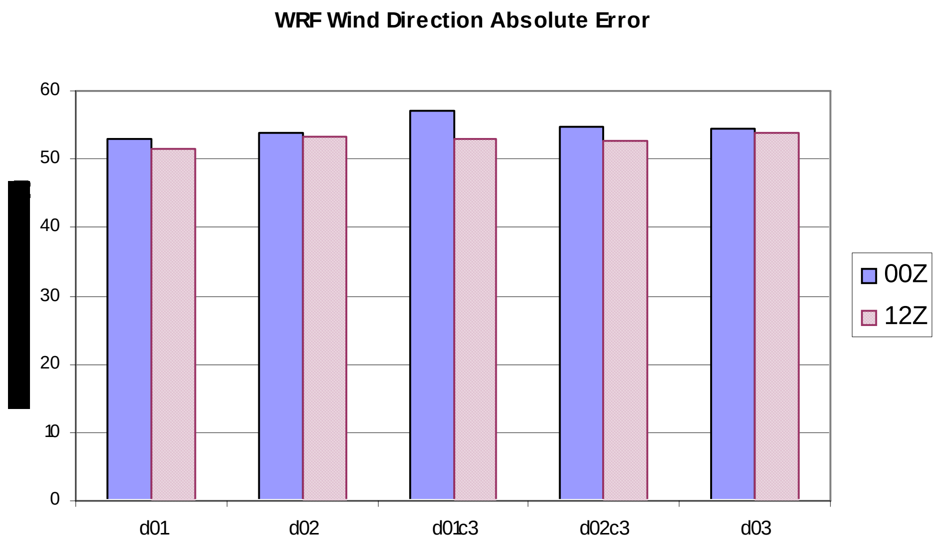


Figure 17. The average absolute wind direction errors for all stations and times for the 00Z and 12Z WRF model runs.

The largest average absolute wind direction error of 57º is for the 00Z 18-km model run cropped to the innermost domain, but that is not significantly different from the other categories, which have average errors ranging between 51º and 55º.

3.1.5 Surface Forecast Error Summary Statistics

Table 4 presents summary statistics for all the surface results in this study. As previously mentioned, the 00Z and 12Z model output covers hourly forecasts from 1 h through 12 h past the model initialization time. The various categories listed are defined in table 3 (under subsection 3.1). Many of the mean difference and absolute difference values are discussed in the previous paragraphs. Table 4 also includes correlation coefficient, root mean square error (RMSE), and root mean square vector error (RMSVE) data.

Table 4. Summary statistics for surface forecast errors.

		00Z Model Runs					12Z Model Runs				
		d01	**d02**	**d01c3**	**d02c3**	**d03**	**d01**	**d02**	**d01c3**	**d02c3**	**d03**
Temperature (°C)	**Correlation Coefficient**	.88	.83	.78	.81	.84	.87	.88	.79	.89	.93
	Mean Error	3.4	3.5	4.1	3.7	3.7	-0.9	-1.1	-0.6	-1.0	-1.0
	Absolute Error	3.9	4.0	4.4	4.0	3.9	3.1	2.8	3.2	2.6	2.3
	RMSE	4.9	5.1	5.6	5.1	5.0	4.0	3.5	4.0	3.2	2.7
	Number of Points	31359	13361	11141	11141	11141	24578	10277	8507	8507	8507
Dew-Point (°C)	**Correlation Coefficient**	.69	.82	.82	.83	.83	.69	.79	.81	.82	.82
	Mean Error	1.4	2.0	2.3	2.2	2.2	0.8	1.1	1.4	1.4	1.4
	Absolute Error	3.6	3.2	3.4	3.3	3.3	3.3	3.1	3.0	3.0	3.0
	RMSE	5.1	4.2	4.4	4.3	4.3	4.8	4.1	4.0	4.0	4.0
	Number of Points	23430	11246	9503	9503	9503	18270	8544	7168	7168	7168
Wind Speed (m/s)	**Correlation Coefficient**	.32	.32	.30	.35	.42	.19	.27	.36	.40	.43
	Mean Error	0.7	0.6	0.0	0.3	0.4	0.3	-0.4	-1.1	-0.8	-0.7
	Absolute Error	2.1	2.3	2.3	2.4	2.3	2.4	2.3	2.4	2.3	2.2
	RMSE	3.1	3.2	3.1	3.1	3.0	3.3	3.2	3.2	3.1	3.0
	Number of Points	17508	11488	9867	9870	9866	13468	8674	7438	7440	7437
Wind Direction (°)	**RMSVE**	5.2	5.7	5.6	5.7	5.8	5.3	5.7	5.6	5.7	5.8
	Absolute Error	53	54	57	55	55	52	53	53	53	54
	Number of Points	17290	11345	9751	9749	9751	13331	8572	7354	7352	7346

3.2 Upper-Air Forecast Evaluations

As described in subsection 2.3, the three upper-air stations used for comparisons to WRF forecasts include Tucson, AZ, (TUS) in the lower Sonoran desert; and Albuquerque, NM, (ABQ); and El Paso, TX, (ELP) in the higher Chihuahuan desert. The TUS and ABQ results are available only for the 18-km grid spacing model output. The ELP results are available for all 3 grid spacings, but since the differences between the 18-km, 6-km, and 2-km results do not appear to be great, the 6-km output is not discussed. To streamline the data handling for the large amount of upper-air data, the upper-air evaluations are limited to forecasts every 5 days throughout the period containing 00Z- and 12Z-initialized forecasts from April 12 through May 15, 2004. For each of the 7 days used, analyses are included for the 0-h and 12-h forecast from both the 00Z and 12Z model run, resulting in 4 upper-air forecasts per day for each location. This approach also lessens the redundancy of adjacent days experiencing similar synoptic conditions.

3.2.1 Forecast Errors by Station at the 50-kPa Pressure Surface

Initially, a single height at 50 kPa for each upper-air forecast evaluation station is used. As described above, each box-whisker plot encompasses 28 forecasts. Figure 18 shows the temperature forecast errors at 50 kPa for TUS, ABQ, and ELP from the 18-km grid-spacing forecast, as well as for ELP from the 2-km forecast.

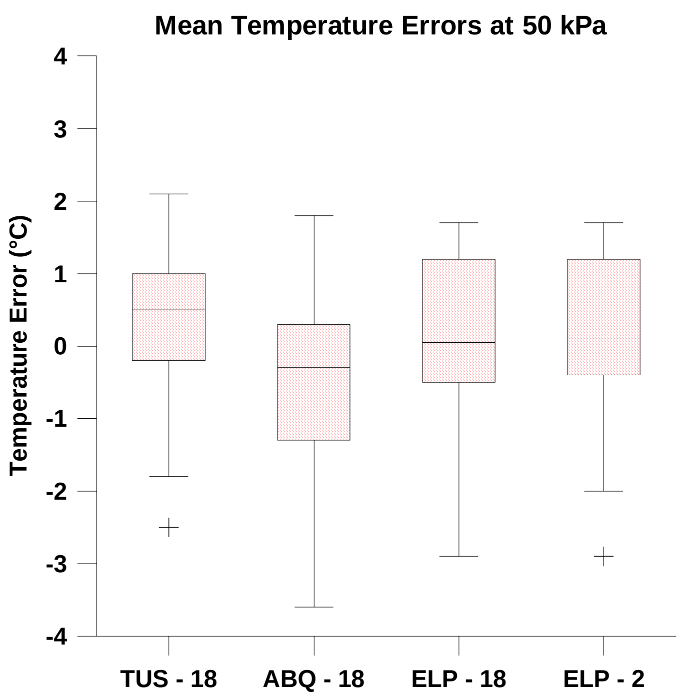


Figure 18. Quartiles of temperature forecast errors at 50 kPa for TUS, ABQ, and ELP based on the 18-km WRF model runs, and for ELP based on the 2-km run.

The 50-kPa temperature forecasts are generally quite good, with half the forecasts within 1.5 ºC or less of the observed value. The most obvious difference between the upper-air stations is the tendency for the ABQ forecast to be too cold, while the other stations' forecasts are more likely to be slightly too warm. The 50-kPa temperature forecasts from the 2-km model runs are either identical or within 0.1 ºC of the equivalent forecast from the 18-km model run.

The dew-point temperature forecasts at 50 kPa were very much worse than the temperature forecasts. Although it is understood that dew-point temperature profiles are significantly more variable than temperature profiles, the errors seen in figure 19 are frequently well outside an acceptable range for operational use. Note that the *y*-axis in figure 19 extends from -30 to 30 ºC compared to -4 to 4 ºC for the temperature errors in figure 18.

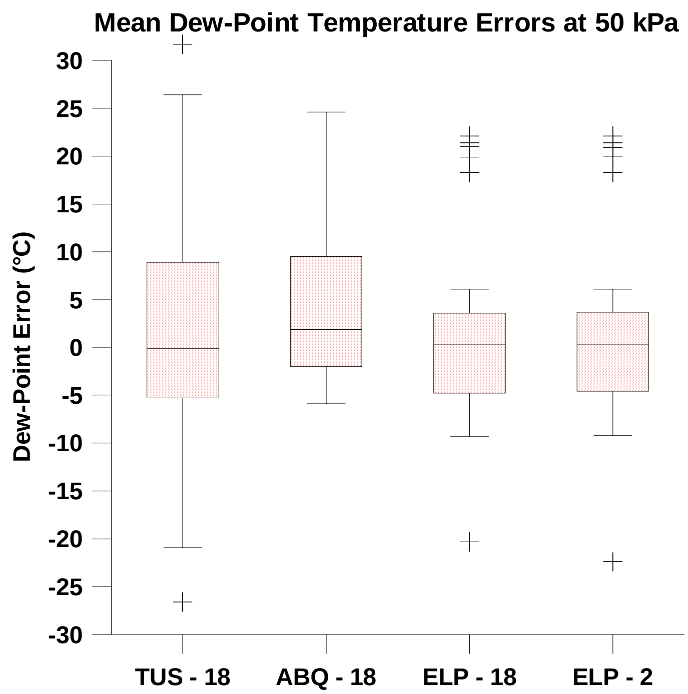


Figure 19. Quartiles of dew-point temperature forecast errors at 50 kPa for TUS, ABQ, and ELP based on the 18-km WRF model runs, and ELP based on the 2-km run.

The IQR for TUS extends from -5 ºC (forecasts are too cold) to 10 ºC (forecasts are too warm), with large errors in either direction common. ABQ experienced similar instances of dew-point temperatures forecasts being substantially too warm. On the other hand, the forecasts that were too cold showed much smaller errors. ELP did not share this warm bias, with the IQR range between -5 and 5 ºC. As with the temperature forecasts, the 18-km and 2-km dew-point temperature forecasts showed no discernable differences.

The wind speed forecasts at 50 kPa tended to be too low, with only the upper quartile containing winds forecast to be too strong (figure 20). The IQR for ABQ encompasses errors between approximately 0 and -10 m/s, while the errors for the other stations are not quite as dispersed.

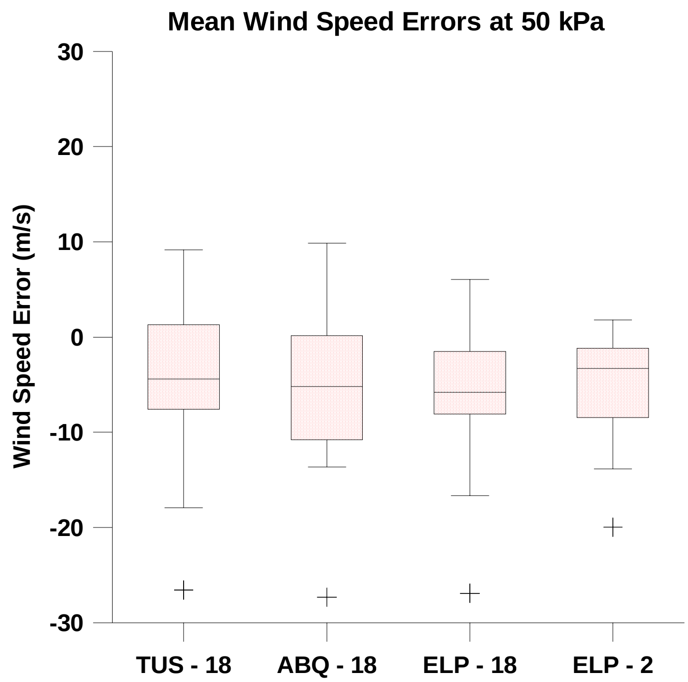


Figure 20. Quartiles of wind speed forecast errors at 50 kPa for TUS, ABQ, and ELP based on the 18-km WRF model runs, and for ELP based on the 2-km run.

Although the IQR is similar for the ELP 18-km and 2-km errors, the 2-km wind speed forecasts are frequently several meters per second different from the 18-km wind speed forecasts and don’t generate the extreme errors seen in the 18-km forecasts.

The 50-kPa wind direction forecasts at TUS and ABQ showed little bias, with half of the forecasts within 40º of the observed value (figure 21). Both the 18-km forecasts and the 2-km forecasts for ELP generated somewhat smaller errors. Although the 2-km wind direction forecasts were frequently 20º or more different than the 18-km forecast, they were not predominantly the better forecast.

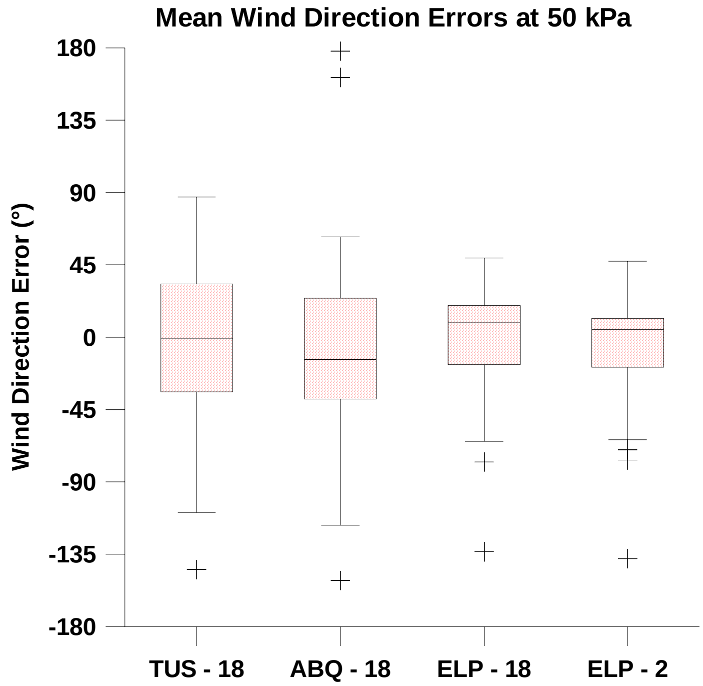


Figure 21. Quartiles of wind direction forecast errors at 50 kPa for TUS, ABQ, and ELP based on the 18-km WRF model runs, and for ELP based on the 2-km run.

3.2.2 Forecast Errors by Time at the 50-kPa Pressure Surface

The same 7 days as used above for the 50-kPa evaluations by station were used to group the average 50-kPa errors by forecast initialization time and forecast hour. The 18-km model forecasts for the 3 upper-air stations are included, resulting in 21 forecasts for each of the following 4 forecast time combinations:

- Forecast initialized at 00Z valid at that time (00Z – 0 h)
- Forecast initialized at 00Z valid 12 h later (00Z – 12 h)
- Forecast initialized at 12Z valid at that time (12Z – 0 h)
- Forecast initialized at 12Z valid 12 h later (12Z – 12 h)

The temperature forecasts at 50 kPa reflected a small bias for being too warm for the 0-h forecasts, and a small bias for being too cool for the 12-h forecasts (figure 22), although the sample size and the small errors overall preclude drawing any conclusions about persistent model bias. As expected, the 12-h forecasts have higher average absolute errors than the 0-h forecasts. The forecasts initialized at 12Z generated somewhat higher errors than those initialized at 00Z for the same forecast hours.

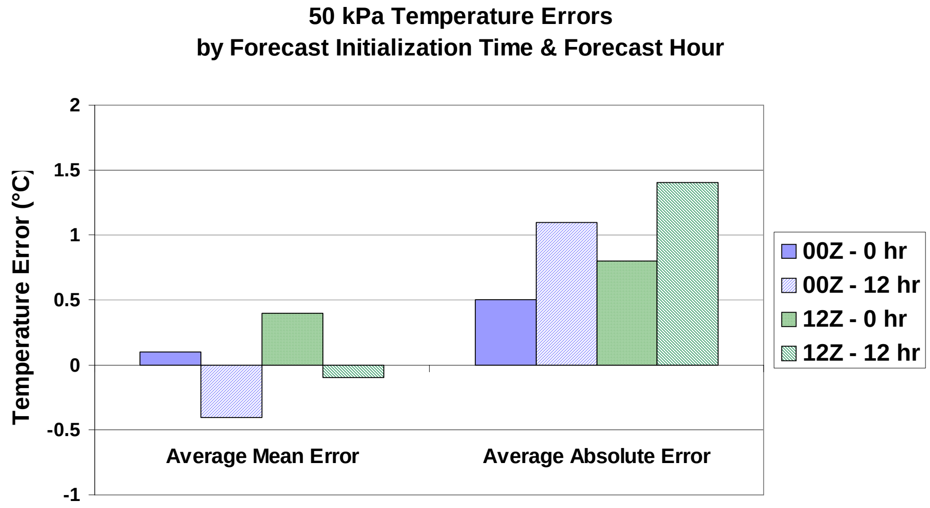


Figure 22. The average mean and absolute temperature errors for the 3 upper-air stations at 50 kPa, broken down by forecast model initialization time (00Z or 12Z) and forecast hour (0 h or 12 h).

As discussed previously concerning the 50-kPa analysis by station, the dew-point temperature forecast errors are significantly greater than the temperature errors. Therefore, the *y*-axis in figure 23 encompasses a much greater range than is used in figure 22. Other differences seen in dew-point temperature errors as compared to temperature errors include a substantial warm bias for the 12-h forecast initialized at 00Z and the greatest absolute error occurring at that time. Seven of the 21 forecasts at that time generated errors greater than 10 °C.

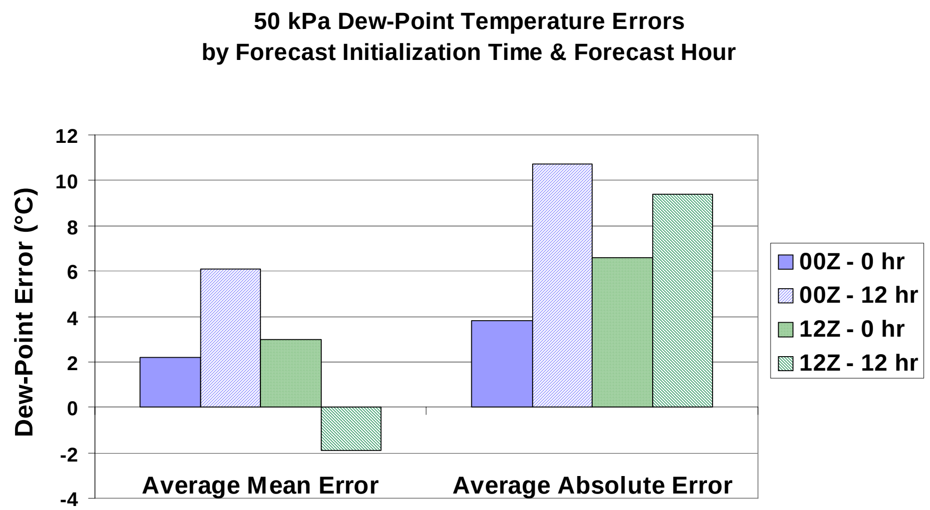


Figure 23. The average mean and absolute dew-point temperature errors for the 3 upper-air stations at 50 kPa, broken down by forecast model initialization time (00Z or 12Z) and forecast hour (0 h or 12 h).

The wind speed forecasts at 50 kPa were biased to be too weak for each of the forecast times (figure 24). The 0-h forecasts initialized at both 00Z and 12Z only show a bias of 2 m/s with an average absolute error of 4–5 m/s. On the other hand, the 12-h forecasts reflect a more predominant bias amount only slightly lower than the average absolute errors, around 7 m/s for the 00Z forecast and 10 m/s for the 12Z forecast.

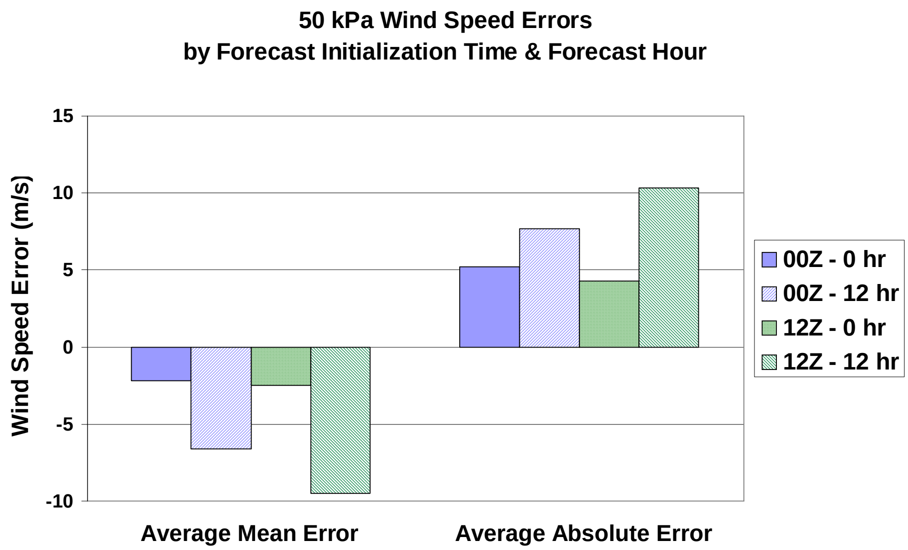


Figure 24. The average mean and absolute wind speed errors for the 3 upper-air stations at 50 kPa, broken down by forecast model initialization time (00Z or 12Z) and forecast hour (0 h or 12 h).

The wind direction forecasts at 50 kPa do not produce significant bias (figure 25). Once again, the 0-h forecasts are significantly better than the 12-h forecasts, with average absolute errors below 20º for the 0-h forecasts and 50º to 60º for the 12-h forecasts.

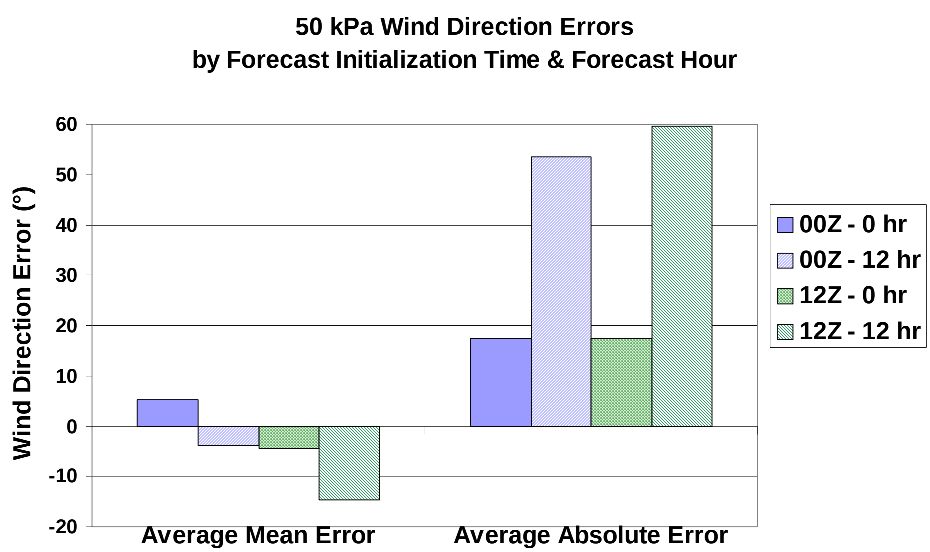


Figure 25. The average mean and absolute wind direction errors for the 3 upper-air stations at 50 kPa, broken down by forecast model initialization time (00Z or 12Z) and forecast hour (0 h or 12 h).

3.2.3 Overall Upper-Air Forecast Analyses

An attempt was made to evaluate the quality of the upper-air forecasts at all sounding heights for the same 7 days included in the 50 kPa discussions above. Based on viewing plots showing both the forecast and the observed upper-air data for the individual weather parameters, a strictly subjective grade A to E was given to each forecast to estimate the quality of the forecast. Each of the 7 days included 4 parameters, 4 station choices (TUS, ABQ, and ELP at an 18-km grid spacing, plus ELP at a 2-km grid spacing) and 4 model initialization and forecast hour choices (00Z initialization with a 0-h and a 12-h forecast and 12Z initialization with a 0-h and a 12-h forecast). Therefore, a total of 448 profiles were examined.

Table 5 gives the summary of how many profiles received a grade of A, B, C, D, or E for each weather parameter grouped by station and by forecast initialization time and forecast hour. The temperature and dew-point temperature plots were virtually indistinguishable between the 18-km and 2-km results for ELP, while the wind plots did show significant differences based on the model grid spacing.

Table 5. Summary subjective grades for upper-air forecast accuracy, by forecast station, forecast initialization time, and forecast hour.

		By Station				By Forecast Time				
		TUS 18 km	ABQ 18 km	ELP 18 km	ELP 2 km	00Z 0 h	00Z 12 h	12Z 0 h	12Z 12 h	Total
Temperature (°C)	**A**	**24**	**23**	**24**	**24**	**28**	**17**	**28**	**22**	**95**
	B	4	5	4	4	0	11	0	6	17
	C	0	0	0	0	0	0	0	0	0
	D	0	0	0	0	0	0	0	0	0
	E	0	0	0	0	0	0	0	0	0
Dew-Point Temperature (°C)	**A**	3	7	7	7	**16**	6	2	0	24
	B	**18**	**16**	**18**	**18**	12	**13**	**25**	**20**	**70**
	C	6	5	2	2	0	9	1	5	15
	D	1	0	1	1	0	0	0	3	3
	E	0	0	0	0	0	0	0	0	0
Wind Speed (m/s)	**A**	1	2	1	4	5	0	3	0	8
	B	**13**	7	4	6	**12**	2	**15**	1	30
	C	5	**13**	**10**	**12**	11	7	10	**9**	**37**
	D	2	1	8	6	0	8	0	**9**	17
	E	7	8	5	0	0	**11**	0	**9**	20
Wind Direction (°)	**A**	4	**9**	1	6	8	0	11	1	20
	B	8	5	11	**9**	**13**	2	**13**	5	33
	C	**9**	5	**13**	**9**	6	**12**	4	**14**	**36**
	D	6	4	2	3	1	8	0	6	15
	E	1	5	1	1	0	6	0	2	8

NOTE: The modal grade for category and for the parameter in total is highlighted in green.

The upper-air temperature forecasts were quite good, with similar results for the different stations. The 0-h forecasts for both initialization times produced very good results, while the 12-h forecasts were frequently not quite as good, especially for the 00Z 12-h forecasts. Eleven of the 28 observed temperature profiles included temperature inversions at the surface for at least one of the stations. Unfortunately the model output did not include data at height levels low enough to investigate whether or not the WRF model accurately predicted the temperature inversions. A typical upper-air temperature result is shown in figure 26.

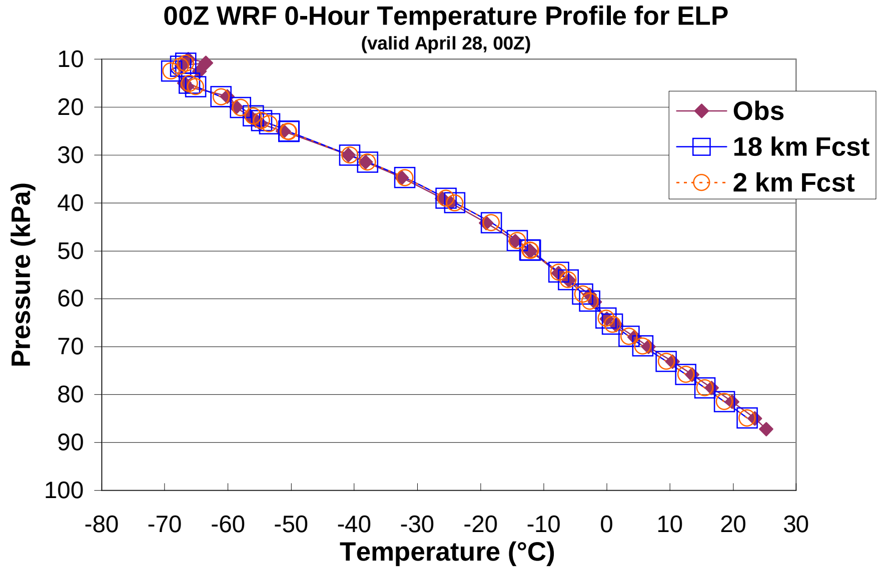


Figure 26. A sample temperature profile rated “A” for a very good quality forecast.

Since observed profiles of dew-point temperatures vary significantly more than temperature profiles, even in a desert environment, it is not surprising that the upper-air dew-point temperature forecasts did not perform quite as well as the temperature forecasts. The three stations predominantly experienced similar results for the same forecast times. Two instances where the TUS forecasts were significantly worse than the other two stations resulted when the TUS observations exhibited a rather abrupt change in the slope of the profile between 60 and 40 kPa, and the WRF did not capture these accurately. The dew-point temperature forecasts produced the best results for 0-h forecasts based on the model runs initialized at 00Z, but the 0-h forecasts based on the model runs initialized at 12Z were also consistently pretty good. The 12-h forecast results were somewhat more diverse, with the worst results associated with the 12-h forecast from the 12Z model runs. Figure 27 shows a representative dew-point temperature profile.

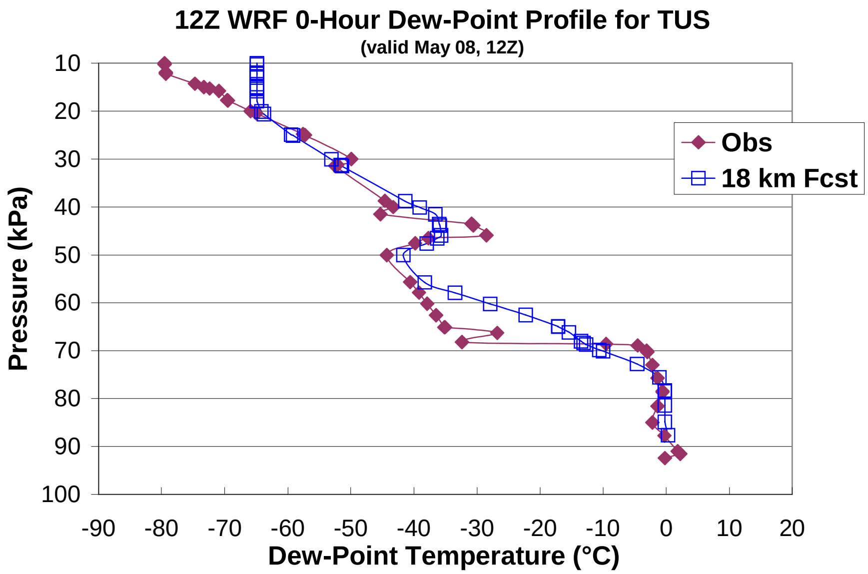


Figure 27. A sample dew-point temperature profile rated "B" for a good quality forecast.

As mentioned previously, the 2-km ELP wind speed forecasts often differed significantly from the 18-km ELP forecasts, with the 2-km forecasts frequently showing an improvement over the 18-km forecasts. Several of the observation times in May 2005 for ELP and ABQ encompassed very high upper-air wind speeds above 40 kPa that were missed by the forecast model. TUS did not report these high wind speeds, so its forecast quality is rated higher more often than the other stations, presumably because of an easier forecast situation. The 0-h wind speed forecasts showed significantly more skill than the 12-h forecasts, which were quite poor overall. A sample wind speed profile is provided in figure 28.

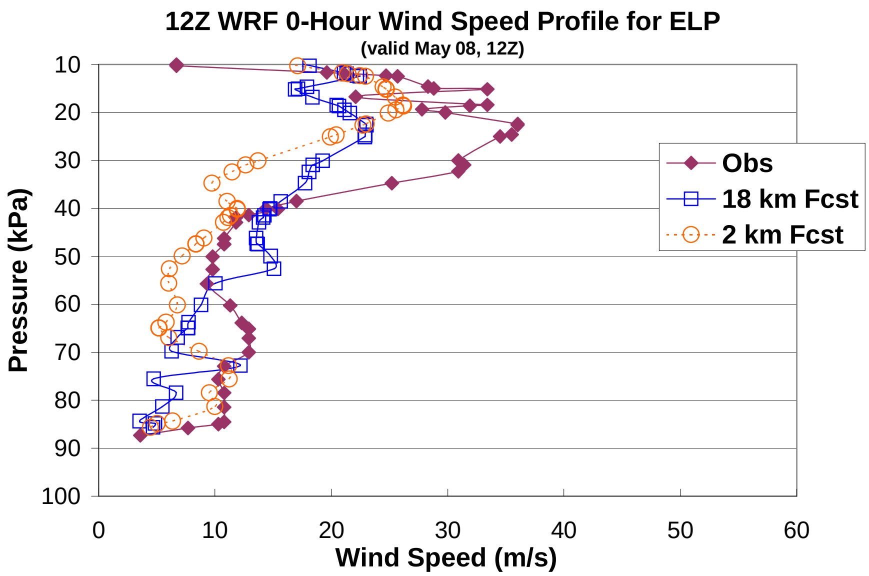


Figure 28. A sample wind speed profile rated "C" for a fair quality forecast.

As found in the wind speed forecasts, there was a wide variety in the quality of the wind direction forecasts. Although the distribution of wind direction forecasts rated B–D was similar to the wind speed distribution, there were many more very good (A) wind direction forecasts and many fewer very poor (E) wind direction forecasts than wind speed forecasts. Interestingly, the ABQ wind direction profiles generated both the highest number of very good and the highest number of very bad forecasts compared to the other stations. The 2-km results for ELP frequently displayed an improvement over the 18-km results for wind direction. The 0-h forecasts were most often considered very good or good, while the 12-h forecasts were predominantly only fair. Figure 29 shows a sample wind direction profile.

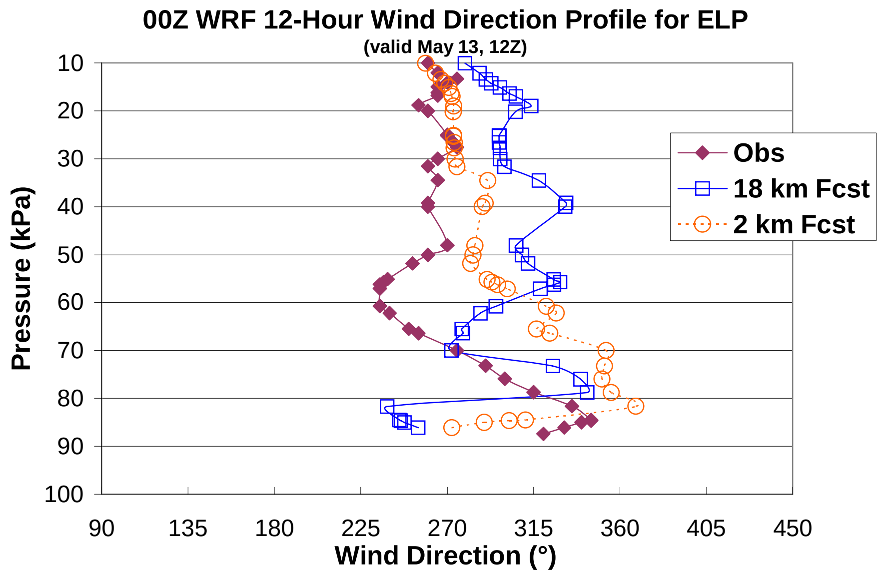


Figure 29. A sample wind direction profile rated “C” for a fair quality forecast.

4. Conclusions

The usual disclaimers apply. One should not draw conclusions from these results concerning the performance of the WRF model over other times and locations. In addition, it must be remembered that some erroneous surface data was likely used in these evaluations, although it is hoped that the amount of bad data was small compared to the total number of data points.

4.1 Surface Forecast Results

Many comparisons were made between WRF forecasts and MADIS and SAMS surface observations. The period considered for this study included only the spring 2005 season, but encompassed many different observation times and locations within New Mexico and Arizona. Each model run only provided output out to 12 h, so the surface statistics for model runs initialized at 00Z cover evening and night times, while those initialized at 12Z cover morning and day times. From the detailed analyses presented in this report, some general conclusions on the WRF forecast performance include the following:

- The temperature forecast errors averaged 2–4 ºC, with a large warm bias of 3 ºC for the model runs initialized at 00Z and a much smaller cool bias of 1 ºC or less for the model runs initialized at 12Z.
- All of the various domain groupings generated average absolute dew-point temperature errors between 3 and 3.5 °C. These included a warm bias for both initialization times. The smallest bias under 1 ºC occurred when considering all the surface stations in domain 1 for the 12Z model runs. The 00Z model runs resulted in 2 ºC or slightly higher biases.
- Wind speed forecasts produced similar average absolute errors between 2 and 2.5 m/s for all the domains and forecast times. The wind speed forecast biases were quite small, but did vary noticeably, from more than 0.5 m/s too strong for the 00Z model runs over domains 1 and 2, to more than 0.5 to 1 m/s too weak for the 12Z model runs for all 3 grid spacings over the area within domain 3.
- The wind direction forecasts experienced the smallest RMSVE of 5.2 and 5.3 m/s when considering all the stations in domain 1. The RMSVE for other domain groupings ranged from 5.6 to 5.8 m/s. The average absolute wind direction errors were approximately 55º.
- Temperature forecasts were slightly better overall for the 2-km model runs than for those with larger grid spacing. The other parameters did not reflect any persistent improvement or decline in the forecast error based on the model grid spacing.

4.2 Upper-Air Forecast Results

A subset of two model runs a day for seven days approximately fives days apart was selected to analyze the upper-air forecast results. A 0-h and a 12-h forecast were included for both the 00Z and 12Z model runs on each of the 7 days. A summary of some of the upper-air forecast results follows:

- The temperature forecast errors at 50 kPa were more likely to be 0–1 ºC too warm at TUS and ELP, and more likely to be 0–1.5 ºC too cool at ABQ. The average absolute 50-kPa temperature forecast error, by time, ranged from 0.5 ºC for the 0-h 00Z forecasts to 1.4 ºC for the 12-h 12Z forecasts.
- TUS and ABQ experienced a greater range of 50-kPa dew-point temperature errors than seen at ELP. The greatest 50-kPa dew-point temperature bias of 6 ºC too warm was associated with the 12-h forecast initialized at 00Z. The biases at the other times were close to 2 ºC, with both 0-h forecasts too warm and the 12-h 12Z forecast too cool. The average absolute 50-kPa dew-point temperature errors were approximately 4 ºC and 6 ºC for the 0-h forecasts initialized at 00Z and 12Z, respectively, and greater than 10 ºC and 9 ºC for the 12-h forecasts.
- Wind speed forecasts at 50 kPa tended to be 0–10 m/s too low for all three stations. The 0-h forecasts were predominantly in the 0–5 m/s half of this range with the 12-h forecasts in the 5–10 m/s half.
- Wind direction forecast errors at 50 kPa fell within 45º in half the cases for TUS and ABQ, with errors frequently half that amount for ELP. The 0-h wind direction average absolute forecast errors were less than 20º for the 0-h forecasts and 50 º to 60º for the 12-h forecasts.

References

1. Michalakes, J.; Dudhia, J.; Gill, D.; Klemp, J.; Skamarock, W. *Design of a Next-Generation Regional Weather Research and Forecast Model: Towards Teracomputing*; World Scientific: River Edge, NJ, 1998, pp 117–124.

2. Black, T.L. The New NMC Mesoscale Eta Model: Description and Forecast Examples. *Wea. Forecasting* **1994,** *9*, 265–278.

3. Mesoscale Branch Web Page Reference List. http://www.emc.ncep.noaa.gov/mmb/mmbpll/eric.html (accessed Dec 2005).

4. Meteorological Assimilation Data Ingest System (MADIS). http://www-sdd.fsl.noaa.gov/MADIS/ (accessed Dec 2005).

5. Goeber, M.; Wilson, C. *Why, when a model resolution is improved, do the forecasts often verify worse?* Australian Bureau of Meteorology. http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/Goeber_Wilson/Double_penalty_combined.html (accessed Dec 2005).

6. Sauter, B. *Forecast Model Uncertainty; MM5 Accuracy Over Utah*; ARL-TR-3170; U.S. Army Research Laboratory: White Sands Missile Range, NM, March 2004.

Acronyms

00Z	0000 Zulu time
01Z	0100 Zulu time
12Z	1200 Zulu time
13Z	1300 Zulu time
ABQ	Albuquerque, NM, upper-air station
ELP	El Paso, TX, upper-air station
FCST	forecast
IQR	Inner Quartile Range
MADIS	Meteorological Assimilation Data Ingest System
MST	Mountain Standard Time
NAM	North American Mesoscale model
NCAR	National Center for Atmospheric Research
OB/OBS	observations
RMSE	Root Mean Square Error
RMSVE	Root Mean Square Vector Error
RRTM	Rapid Radiation Transfer Model
SAMS	Surface Atmospheric Measuring System
TUS	Tucson, AZ, upper-air station
WRF	Weather Research and Forecasting model
YSU	Yonsei University

Distribution List

	Copies
HEADQUARTERS DEPT OF ARMY DAMI POB WEATHER TEAM 1000 ARMY PENTAGON ROOM 2E383 WASHINGTON DC 20310-1067	1
US ARMY RESEARCH LAB AMSRD ARL D 2800 POWDER MILL ROAD ADELPHI MD 20783	1
US ARMY RESEARCH LAB AMSRD ARL CI E BATTLEFIELD ENVIR DIV ADELPHI MD 20783-1197	1
US ARMY RESEARCH LAB AMSRD ARL CI EM BATTLEFIELD ENVIR DIV WSMR NM 88002-5501	1
US ARMY RSRC OFC ATTN AMSRD ARL RO E DR BACH PO BOX 12211 RTP NC 27009	1
US ARMY DUGWAY PROVING GROUND STEDP MT DUGWAY UT 84022	1
US ARMY FIELD ARTILLERY SCHOOL ATSF TSM TA FT SILL OK 73503-5600	1
US ARMY TOPO ENGR CTR CETEC ZC 1 FT BELVOIR VA 22060-5546	1
WHITE SANDS FORECAST SECTION CSTE DTC WS RO DA F WSMR NM 88002	1
WSMR TECH LIBRARY BR STEWS IM IT WSMR NM 88002	1
HQ AFWA/DNX 106 PEACEKEEPER DR STE 2N3 OFFUTT AFB NE 68113-4039	1

	Copies
BOB CRAIG HQ AFWA 106 PEACEKEEPER DR STE 2N3 OFFUTT AFB NE 68113-4039	1
AFRL TECH LIB 29 RANDOLPH RD HANSCOM AIR FORCE BASE MA 01731-3010	1
AFCCC/DOR TECH LIB 151 PATTON AVE RM 120 ASHEVILLE NC 28801-5002	1
NAVAL RESEARCH LAB MARINE METEOROLOGY DIVISION 7 GRACE HOPPER AVENUE STOP 2 MONTEREY CA 93943-5502	1
JEFFREY A LERNER PHD MODELS TEAM MANAGER FNMOC 7 GRACE HOPPER AVENUE MONTEREY CA 93943	1
MARK SWENSON FNMOC 7 GRACE HOPPER AVENUE MONTEREY CA 93943	1
NCAR LIBRARY SERIALS NATL CTR FOR ATMOS RSCH PO BOX 3000 BOULDER CO 80307-3000	1
CHRIS DAVIS NCAR/MMM PO BOX 3000 BOULDER CO 80307-3000	1
STAN TRIER NCAR/MMM PO BOX 3000 BOULDER CO 80307-3000	1
JENNIFER MAHONEY NOAA/ESRL R/GSD5 325 BROADWAY BOULDER CO 80305-3328	1

	Copies
YING LIN NOAA/NCEP ENVIRONMENTAL MODELING CENTER 5200 AUTH ROAD CAMP SPRINGS MD 20746-4304	1
FRED CARR OU/CAPS SARKEYS ENERGY CENTER STE 1110 100 EAST BOYD STREET NORMAN OK 73019	1
US ARMY RESEARCH LAB AMSRD ARL CI EM B SAUTER WSMR NM 88002	3 hard copies
US ARMY RESEARCH LAB ATTN IMNE AD IM DR MAIL & RECORDS MGMT ADELPHI MD 20783-1197	1

	Copies
ADMNSTR DEFNS TECHL INFO CTR ATTN DTIC OCP 8725 JOHN J KINGMAN RD STE 0944 FT BELVOIR VA 22060-6218	1
US ARMY RESEARCH LAB AMSRD ARL CI OK TL 2800 POWDER MILL ROAD ADELPHI MD 20783-1197	2 hard copies
US ARMY RSRCH LAB ATTN AMSRD ARL CI OK TP TECH LIB T LANDFRIED APG MD 21005-5066	2
TOTAL	32 (27 CDs and 5 hard copies)